



An ontology-based method for quality assessment of spatial data bases

Mir-Abolfazl Mostafavi, Geoffrey Edwards, Robert Jeansoulin

► To cite this version:

Mir-Abolfazl Mostafavi, Geoffrey Edwards, Robert Jeansoulin. An ontology-based method for quality assessment of spatial data bases. Third International Symposium on Spatial Data Quality, Department for Geoinformation and Cartography, Vienna University of Technology, Apr 2004, Bruck an der Leitha, pp.49-66. inria-00000447

HAL Id: inria-00000447

<https://inria.hal.science/inria-00000447>

Submitted on 17 Oct 2005

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

AN ONTOLOGY-BASED METHOD FOR QUALITY ASSESMENT OF SPATIAL DATA BASES

Mir Abolfazl Mostafavi¹, Geoffrey Edwards², Robert Jeansoulin³

1: mir-abolfazl.mostafavi.1@ulaval.ca

Centre de recherche en géomatique
Université Laval, Québec, Canada

2: geoffrey.edwards@scg.ulaval.ca

Canada Research Chair in Cognitive Geomatics
Centre de recherche en géomatique
Université Laval, Québec, Canada

3: jeansoul@gyptis.univ-mrs.fr

Laboratoire d'Informatique de Marseille,
Université de Provence, Marseille, France

Abstract

This paper proposes an ontological approach for quality assessment of the spatial databases. This process is carried out at two levels. At the ontological level, the internal consistency of the specifications is considered. At the data level, real objects and their relations are studied with respect to the specifications. For this purpose, the national topographic database of Canada is selected as the case study. The ontology of the spatial database is translated into a knowledge base coded in Prolog. Then rules that define inconsistencies were defined. The querying of the knowledge base to determine the existence of such inconsistencies was carried out on a very large fact base. By this process, spatial relation between each pair of objects is analyzed with respect to the permitted relation between such objects in the ontology. The results obtained from various experimentations indicate the presence of several inconsistencies in the analyzed data set. These problems were attributed firstly to the control system that oversees the production process and secondly to the incomplete ontology. The overall approach appeared to be justified. The results obtained from several experimentation illustrated the potential of the proposed method for the quality assessment of spatial data bases in both ontological and data levels.

Key Words

Logical consistency, quality, ontology, spatial data bases, Prolog, AI.

1. Introduction

The widespread use of spatial data for decision making, navigation, planning and development, natural resource management, and emergency services, the increasingly questionable quality of data produced for both public and private sectors and, also, the easy access to secondary data sources following the growth of the internet and other communication technologies all raise concerns about the quality of spatial data (Goodchild and Jeansoulin 1998; Guptill and Marrison 1995; Shi et al. 2002). In order to reduce the risks associated with using such information, we need formal methods to evaluate the quality of the data. Quality depends on several factors; the 'internal' quality of spatial data is determined by its actuality, geometric and semantic accuracy, its genealogy, logical consistency and the completeness of the data. This view reflects the producer's perception of quality, which differs from the notion of 'external quality' where the focus is on 'fitness for use'. In the latter definition, quality is defined as the level of fitness between the data and the needs of users. This is generally characterized by elements such as the accessibility, relevancy, completeness, timeliness, interpretability, the ease of understanding and the costs of the data. Based on these definitions, various assessment methods have been proposed for each aspect of data quality (Verigin and Hargitai 1995). These methods are usually based on locally defined product specifications (e. g. NTDB, the National Topographic Data Base of Canada) or pre-defined standards for spatial data such as those of ISO and SDTS.

The specifications published by the producers may be incomplete and inconsistent and as a result the data produced based on these specifications may not satisfy the user's needs. To ensure quality, we need to formally specify appropriate concepts and the relations between these. This implies that we need to refer to the ontology of the spatial data ((Guarino and Giaretta 1995)). "Ontology" is defined as that branch of philosophy which deals with the nature and the organization of reality. In the geospatial data domain, ontology deals with the totality of geospatial concepts, categories, relations and processes and with their interrelations at different resolutions (Guarino and Giaretta 1995; Mark et al. 2001; Smith and Mark 2001). Hence the study of ontology is fundamental to the evaluation of the quality of spatial data, information exchange and interoperability (Egenhofer and Herring 1990).

The logical consistency of a spatial database constitutes an important part of the determination of the internal spatial data quality. It may be defined as the degree of consistency of the data with respect to its specifications. Integrity constraints are the major components of internal consistency. These constraints are often expressed as rules within the specifications. Different forms of rules may be present - explicit rules in the specification itself which concern either the semantics or the geometry of objects, and implicit rules (e.g. a river always flows downstream). However the diversity of the rules and the terminologies that concern geographic data complicate spatial data sharing, reuse, and spatial data fusion. The study of ontology may contribute to the unification of these different conceptualizations.

The remainder of this paper is organized as follows. Section 2 describes the proposed methodology. Section 3 introduces the product ontology as well as the taxonomy of the

NTDB and determines the mapping of the spatial relations specified in the database to the well-known binary topological relations developed by Egenhofer and his collaborators (Egenhofer 1993; Egenhofer and Franzosa 1991; Egenhofer and Herring 1990). Section 4 describes the formalization of the ontology using the Prolog. Section 5 discusses the results concerning inconsistencies in the ontology. Section 6 presents a discussion of the relationship between the consistency of the data and its ontology. Finally, an overall evaluation of the ontology of the NTDB and conclusions are presented in Section 7.

2. Artificial intelligence and quality assessment for spatial databases

The method proposed for evaluating the internal consistency of the database begins with the translation of the data base ontology into a first-order predicate representation. The translation results in a knowledge base that represents the ontology in a formal way. The knowledge base is then completed in Prolog by adding rules designed to test for inconsistencies. Indeed, consistency cannot be tested for directly in Prolog. Instead, inconsistency, when coded as a set of rules, can be tested. Hence the general approach adopted consists of introducing rules for characterizing inconsistency.

In a second stage, the logical consistency of the data set with respect to this ontology is considered. The methodology consists in analyzing the spatial relations within a data set using a commercial GIS and expressing the results in terms of the relations supported by the data base ontology. This requires that one obtain the mapping between spatial operations in the GIS and the relations allowed in the product ontology. If this is done correctly, any binary relation in the data set can be retrieved and represented in propositional form. The formalization can, as a consequence, still be carried out within Prolog.

In a final stage, the two knowledge bases (one for the ontological level and one for the data level) are integrated into a new, combined knowledge base. Rules for inconsistency are defined for both the ontological and the data levels and then queries are established. Finally, results are validated with respect to existing information from other sources. The overall process is presented in Figure 1.

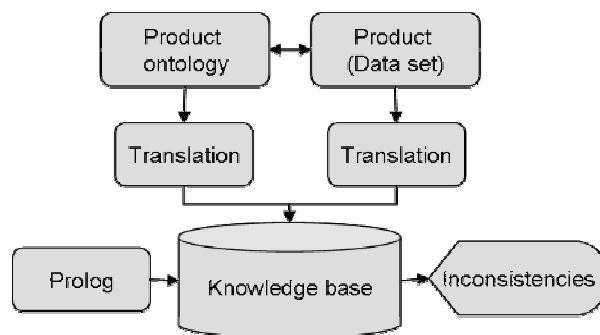


Figure 1 Consistency study for spatial data bases.

3. The Product specifications

Map specifications provide the foundations for product ontologies. Data base ontology may be subdivided into product ontology and user ontology. Fitness for use will not be achievable if the two ontologies do not match. Consistency and completeness are two important elements for this purpose which should be considered in any product ontology.

As a case study for this research, the ontology of the National Topographic Data Base (NTDB) of Canada was considered. This ontology is formulated in two parts within the formal specification provided for the NTDB: one containing entity definitions and their relations (e.g. connection, and sharing), and the other an appendix containing several tables (A to G). Table A describes the entities, combinations of attribute values, NTDB codes and the themes. Table B contains the bilingual information of entities and combinations of attribute values. Table gives the information on the connection relations and the cardinalities of the connection. Table gives the information about the sharing relations and the cardinalities. In formation on the adjacency and superposition relations are given in table E. Minimal allowed dimensions and orientation relations is given in table F and the details about the allowed attributes are given in table G.

In the following two subsections, the taxonomy of the NTDB and the spatial relations that are modeled in it are presented. These together form the principle elements of the product ontology.

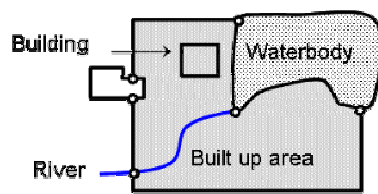
3.1 The taxonomy of the NTDB

The taxonomy of the NTDB largely corresponds to that of the topographic maps in the National Topographic System of Canada (NTS). The territory is the base unit used to manage the data. Each territory is identified by an NTS number, a name and finally by its geographic location (UTM zone). For each territory there is at least one data set. Each data set may contain the topographic information related to different themes (see Table 1). Metadata such as the planimetric and altimetric reference systems used, measurement units and intervals for contours are also given for each data set.

The NTDB embraces 112 types of entity (CTIS, 2002). An entity is defined as a thing that can be distinguished from its surroundings (NTDB Data Dictionary). In the NTDB, an entity is the digital representation of all or part of a topographic feature. Each type of entity is identified by a name (e.g., road, transmission line, lake, etc.) and can be represented by a point, line or a polygon. The geometric representation of an entity may also be characterized by a qualifier that identifies the nature of the geometric representation in terms of its position. This qualifier may take on different values, including its definite position, approximate position or virtual position (e.g. the neatline). This qualifier is defined for point and line representations but not for polygons. This is because a polygon can be a combination of several polylines and each polyline may have its own geometric qualifier. This information is unique to the NTDB and is not usually considered by existing GIS.

3.2 Spatial relations within the NTDB

To ensure the spatial integrity of data within the NTDB, four relations are defined. These relations are connection, sharing, adjacency and superposition (Figure 2). Entity occurrences are *connected* if there is a geometric intersection of the entity occurrences involved. *Sharing* exists between entities when the occurrences of linear or polygonal entities are partially or totally contiguous or coincident. Entity occurrences are *adjacent* when the occurrences are spatially contiguous. Finally, entity occurrences are *superimposed* if they share part or all of their internal regions. Each of these spatial relations must be supported by the NTDB and must be explicitly mentioned in the NTDB specification.



Entity combination	Relation	Entity combination
Built up area	Adjacent	Building
Built up area	Superposed	Building
Built up area	Share	Water body
Watercourse	Connected	Water body

Figure 2 Spatial relations in the NTDB

Cardinalities play an important role in the definition of connection and sharing relations. For the connection relation, three cardinalities are defined. The first type of cardinality (C1) indicates the minimum and maximum extrema of a line segment involved in a connection relation (e.g. C1=0 means no connection is allowed). The second type of cardinality (C2) defines the minimum and maximum number of connections between two occurrences of the same entity. And finally, the third type of cardinality (C3) indicates the minimum and maximum number of connection relations of an occurrence of an entity with respect to occurrences of other entities. The same set of cardinalities is also defined for the sharing relation (see Section 3.1).

Adjacency and superposition relations are essentially defined for polygonal features. In the case when two objects share a common surface area, there is a superposition relation, whereas when two objects share a common border but not a common surface then they are adjacent. Using the object semantics, a superposition relation can be forbidden or allowed. For example the superposition of a highway service center with a built up area is illegal while a hospital can be superimposed on a built-up area.

According to the NTDB specification, additional conditions must be met to support the relations defined above. For the connection relation, if the objects are from two different themes, an intersection point must be added to the geometric representation of the objects involved. In the case where the objects are from the same theme, the objects should be segmented at the connection point (Figure 3). For the sharing relation, if the objects are from different themes, the start and end points of the common part must be added to the

geometry of the objects involved and in the case where the objects are from the different themes, segmentation must be done for the common part.

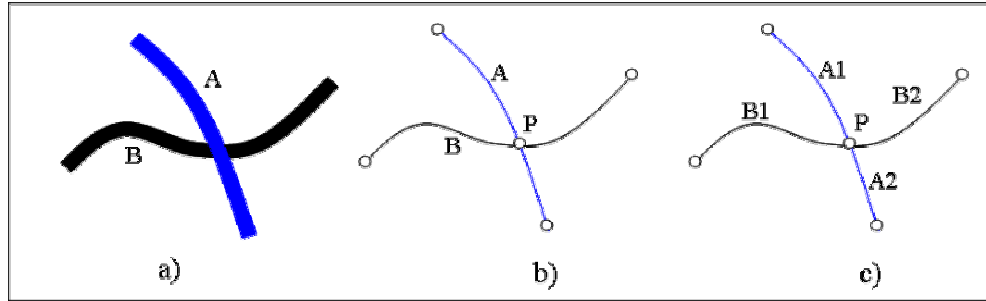


Figure 3 Connection relation a) real objects b) objects from different themes, c) objects from the same theme.

To proceed with the evaluation of consistency using the logical approach, there is a need to determine the compatibility of the data with regard to its specifications, and to carry out this operation within a GIS. Hence, the spatial relations in the NTDB must be mapped to the more common categorization of topological relations. The 9-intersection model defined by Egenhofer and Herring (1990) and Egenhofer and Franzosa (1991) has become standard. In this model, eight types of relation are defined: disjoint, meets, equals, inside, contains, covered by, covers and overlap. In the 9-intersection model, for three elementary objects (point, line, area), there are six groups of candidate relations, including area/area, area/line, area/point, line/line, line/point, point/point relations.

For this project, we have used the Geomedia GIS software. Here, the topological operations are slightly different from the nine intersection relations. The “Meet”, “Overlap”, “Contains”, “Contained by”, “Entirely Contains”, “Entirely Contained by”, and “Spatially Equal” defined in the Geomedia GIS correspond to the “Meet”, “Overlap”, “Covers”, “Covered by”, “Contains”, “Contained by”, and “Equal” relations in the 9-intersection model respectively. The “Touch” relation in Geomedia denotes any kinds of intersection relation between objects. In other words, the “Touch” relation is the opposite of the “Disjoint” relation in the 9-intersection model.

We have expressed the NTDB relations (“Connection”, “Sharing”, “Adjacency” and “Superposition”) in terms of these the Geomedia relation. Table 1 shows an example of the mapping between the “Meet” relation for different class of objects (points, lines and polygons) and the NTDB relations.




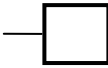

Meet in Geomedia					
Relation					
	Line meets point	Line meets line	Point meets polygon	Line meets polygon	Polygon meets polygon
Connection	x	x	x	x	-
Sharing	-	-	-	-	x
Adjacency	-	-	-	-	x
superposition	-	-	-	-	-

Table 1. Meet relation in Geomedia and its correspondence in the NTDB.

Within a logic system, all statements must be consistent and complete. These fundamental conditions must be satisfied for the mapping of the spatial relations between two different spatial environments. To reproduce the NTDB relations, only the “Touch” and “Overlap” Geomedia operations are needed. Hence inconsistent relations between objects are defined based on the relation found between any pair of objects in the map using the Geomedia operations and the permitted relations within the NTDB ontology. Table 2 shows the Prolog rules used to define inconsistent relations at the data level.

Inconsistent relation	Prolog rules
Inconsistent connection relations	$\text{inconsistent}(a,b) :- \text{orcCard1}(a,b) \wedge \text{dtouch}(a,b)$ $\text{inconsistent}(a,b) :- \text{orcCard2}(a,b) \wedge \neg \text{dtouch}(a,b)$ $\text{inconsistent}(a,b) :- \text{orcCard2}(a,b) \wedge \text{dtouch}(a,b) \wedge \text{dsametheme}(a,b) \wedge \neg \text{segmented}(a,b)$ $\text{inconsistent}(a,b) :- \text{orcc2}(a,b) \wedge \text{dtouch}(a,b) \wedge \neg \text{dcommonpoint}(a,b)$
Inconsistent sharing relations	$\text{incoherent}(a,b) :- \text{orsCard1}(a,b) \wedge \text{dtouch}(a,b)$ $\text{incoherent}(a,b) :- \text{orsCard2}(a,b) \wedge \neg \text{dtouch}(a,b)$ $\text{incoherent}(a,b) :- \text{orsCard2}(a,b) \wedge \text{dtouch}(a,b) \wedge \neg \text{dcommonline}(a,b)$
Inconsistent superposition / adjacency relations	$\text{incoherent}(a,b) :- \text{orsuCard1}(a,b) \wedge \text{doverlap}(a,b)$ $\text{incoherent}(a,b) :- \text{orsuCard2}(a,b) \wedge \neg \text{doverlap}(a,b)$

Table 2. Inconsistent relations at the data level where relations starting with the letter ‘o’ denote the ontological level and relations starting with the letter ‘d’ denote the data level.

4. Formalizing the ontology

In order to evaluate the ontological consistency of the NTDB, first the specifications must be formally represented in a logic framework. For this purpose, Prolog is used. Prolog is a declarative programming language for symbolic non-numeric computation (Bratko

2000). It was specifically developed for solving problems that involve objects and relations between objects. A Prolog program is a collection of two types of logical formulae: facts (or unit clauses) and rules (Head :- Body). Rules and facts will form a *knowledge base* which is used to perform inferences. To initiate computations in Prolog, one needs to query the knowledge base. Processing occurs by backward chaining.

The ontology of the NTDB has more than 300 000 elements. Manual translation of such information is extremely time-consuming. In order to help the process of translating these facts into a logic-based formulation, an interface using the Delphi environment was developed (Mostafavi 2002). This tool ensures the automatic translation of the NTDB ontology into Prolog clauses. The tool uses both the taxonomy information as well as the information concerning the allowed spatial relations, and translates them into Prolog clauses. The translated facts are stored in text files and then transferred to the Prolog environment.

Once formulated in Prolog, the knowledge bases for the taxonomy and the spatial relations were first examined independently. Following this, the knowledge bases were fused and new rules and queries developed to test the internal consistency. In the following section several examples of the facts and rules are discussed.

4.1 Formal representation of the taxonomy and spatial relations

The taxonomy of the NTDB is comprised of several kinds of information. As indicated earlier, each entity has an NTDB code and belongs to a specified theme. A “combination” attribute provides information on the fixed attributes of each entity within a limited domain of authorized values. For example, the “building” entity has 41 fixed attributes including hospital, museum, arena, etc. The information concerning the geometric representations of these entities is also present in the taxonomy. The point representation of an entity has the same code as the NTDB code, while line and surface representations of an entity use a modified code (the NTDB code + 2048 for lines and the NTDB code +4096 for polygons). The geometric representations may vary across scale (e.g. between 1:50000 and 1:250000). A generic clause in Prolog that formalizes this information is as follows:

```
1. taxon(Theme, Entity, Combination, Code, Type, Point50k, Line50K, Surface50K,
Point250k, Line250K, Surface250K).
```

A second generic clause formalizes the connection relation between two objects in the NTDB. This clause is written as follows.

```
2. orrelation (connection, Code1, Attribute, Cardinality1, Cardinality2, Cardinality3, LAT,
LON, list (Code1, Cardinality1, X)) :-list(Code1, Cardinality1, X);
```

where “code1” represents an entity that has a connection relation with a list of entities (X). The “attribute” in this clause indicates the geometric representation qualifier (see Section 2.1). The cardinalities are defined for the connection relations with following values:

- *Cardinality 1*: indicates the number of the extrema of a line segment in a connection relation. Different values of this cardinality are: “0,0” for no connection; “1,1” for connection at only one end; “1,2” for obligatory connection at one end and possible connection at the other end; and finally “2,2” for obligatory connection at both ends of the line segment.
- *Cardinality 2*: indicates the minimum and maximum number of connections between two occurrences of the same entity. The value of the cardinality “1,1” indicates only one connection between the occurrences of the same entity, while “-:” indicates that the nature of the relation is not identified.
- *Cardinality 3*: indicates the minimum and maximum number of connection relations that an occurrence of an entity may have at one of its extrema with occurrences of other entities. The values of this cardinality are “1,1” for one obligatory connection, “1,2” for one obligatory connection and two possible connections, “2,2” for two obligatory connections, “2,*” for two obligatory connections and possible n connections, and “-:” for connection relations between two occurrence of objects that are not identified.

In the Prolog clause presented above, “LAT” indicates the lateral tolerance (e.g. 3m) and “LON” the longitudinal tolerance (e.g. 10m) for the connection relation between two objects. This implies that if the lateral distance between two objects is less than the specified tolerance or if the distance in the direction of an object with regard to another object is less than 10 meters then there will be a connection relation between the specified objects. If the connection relation between two objects is illegal, then the objects should be displaced to satisfy the specified tolerance. The “list” variable in the Prolog clause refers to all of the objects having connection relations with the object identified by “Code1”.

The next generic clause formalizes the sharing relation in the NTDB:

```
3. orelation (sharing, Code, Type, Cardinality1, Cardinality2, Cardinality3, LAT, LON,
list(Code1, Cardinality1, X)):- list(Code1, Cardinality1, X).
```

Three additional types of cardinality are defined for this type of relation, distinct from those defined earlier. Finally the clause that formalizes both adjacency and superposition relations is as follows:

```
4. orelation(adjacency, list( M, X), Attribute, Relationship, Constrain, list( N, Y),nil):- list(M,
X), list( N, Y).
```

This clause identifies the adjacency/superposition relations between two lists of objects (X, Y). If superposition is illegal then the objects are considered to be adjacent. These topological relations may be constrained by an attribute for a given object. For polygonal objects, the superposition relation may include only the interior of the object or its borders as well. This is identified by the values of “Relationship” in the clause as “superposition” and “#superposition”.

This completes the translation of the topological part of the ontology of the NTDB into Prolog clauses. Table 3 shows the total number of facts for each class of relation.

Type of Facts	Groups	Facts
<i>Taxon</i>	368	386
<i>Connection</i>	574	330 796
<i>Sharing</i>	523	15 853
<i>Adjacency/superposition</i>	138	1637
<i>Total</i>		348672

Table 3. Total number of facts in the NTDB

5. Spatial data consistency

Once the ontology of the NTDB is formalized, the logical consistency of a data set can be examined. Logical consistency describes the fidelity of relationships encoded in the data set. The consistency may be studied with respect to two main factors; geometric and semantic constraints. Geometric constraints consist of all rules that should be respected in order to have a clean geometric file. These constraints are explicitly defined in the ontology and are formalized in previous section. Semantic constraints depend on the definition of objects. A relation may be allowed or be forbidden between two objects with respect to their semantics (ex. In the NTDB, the connection relation between an oil and gas facility and a building is forbidden). Both the geometric and the semantic constraints defined by the ontology of the NTDB are formalized in the previous sections and will be used for consistency evaluations of the data from NTDB.

The process of evaluating the logical consistency of the data is comprised of several steps. First, the data set is analyzed within a commercial GIS (ex. Geomedia GIS) and spatial relations between each pair of objects are retrieved using the operations defined in the GIS. These relations then are expressed (see section 3.1) in terms of the NTDB relations. Following this, they are represented formally in Prolog and integrated with the existing knowledge base. Finally, logical inconsistency rules and queries are defined.

For this experiment, the 21E05 file from the NTDB covering most of the Sherbrooke area is considered. In these file 68 different entity classes are represented in separate layers. These layers, which are stored in a Geomedia warehouse, contain over 23000 objects with different geometric representations. For a complete analysis of the binary relations of this file 23000^2 relations must be examined. This represents an enormous amount of computing time. Hence, only a subset of the relations is considered for this study.

For this purpose a spatial analysis module in Visual Basic (VB) was developed. The VB module operates directly with Geomedia. It takes two classes of objects as inputs and then analyses the binary spatial relations between all the objects within these layers. Following this, the relations are expressed in terms of the NTDB spatial relations and are

translated to Prolog clauses. Here are some examples of such formal relations obtained from the VB module:

- drelation(sharing,9,built-uparea,4282,co,6023,road,3649,rr).
- drelation(connection, 245,road,3659,rr, 293,bridge,2142,co).

The first relation indicates that the built-up area #9 from the construction theme shares its borders with the road #6023 of the road network. The second statement expresses a connection relation between the road #245 and the bridge #2142.

Considering the definition of ‘logical consistency’, an inconsistent relation is a relation that exists in the data file but is forbidden in the product ontology, or a relation that is obligatory in the product ontology and is not respected in the data set. These are the basic rules that compare the data to its ontology. The following is an example of a rule that defines the inconsistent connection relation between two objects within the data set. These two objects are identified by Code1 and Code2 which correspond to their respective NTDB identification codes.

- inconsistent (A,B):- orelation(Connected,Code1, A, B, C, D, E, F, list(Code1, Card1, X)), belongsTo(Code2,X), =(Card1, 0,0), drelation(Connected, Code1, Code2).

where “orelation” and “drelation” denote the relations between Code1 and Code2 at the ontological level (connection is forbidden in the ontology because of the definition of the cardinality C1) and at the data levels respectively (connected in data set).

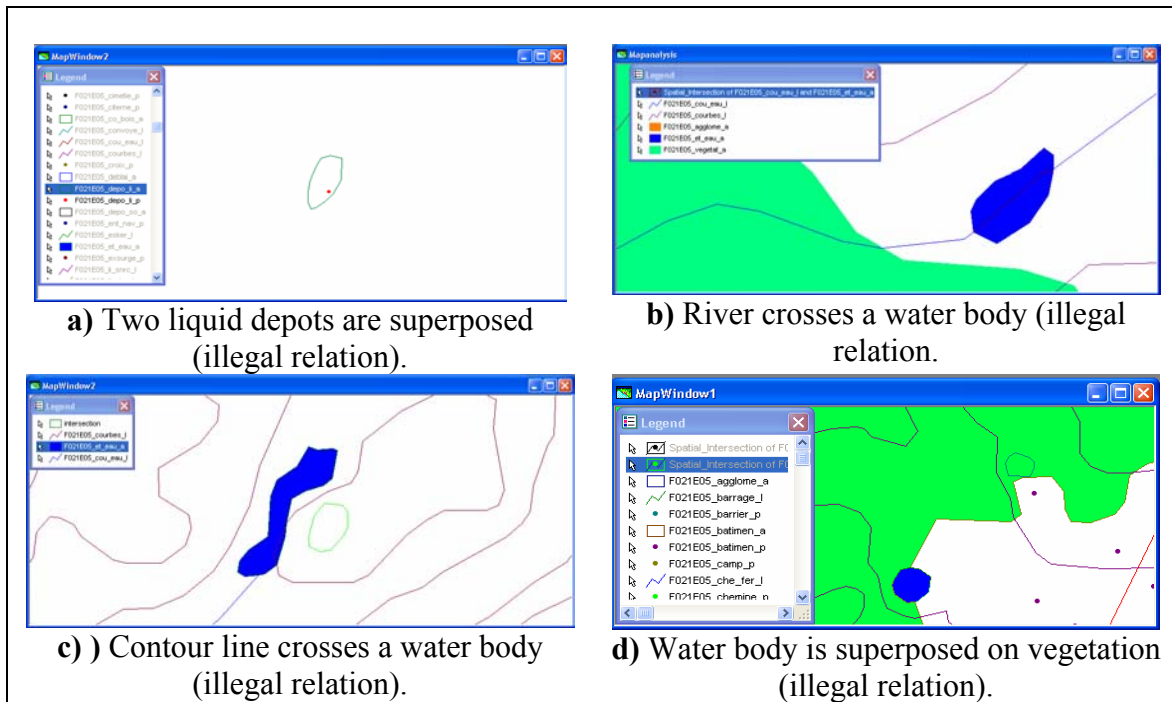


Figure 4 Examples of illegal superposition relations in the NTDB data

Figure 4 illustrates four examples of inconsistent relations in the selected file. Figures 4a and 4b show two superposition relations. These relations are explicitly forbidden in the product ontology but are present at the data level. In contrast to these relations are the “superposition” relations in Figures 4c, and 4d. These rules are not expressed explicitly in the ontology of the NTDB, and hence we can conclude that these problems are due to an incomplete ontology.

The Center for Topographic Information at Sherbrooke (CITS) has defined a criterion to accept or refuse a data file produced by a private company. The criterion defines the percentage of the errors that is tolerated for each data file. For example, only 1% of errors are tolerated for the connection relation whereas 2% of errors are accepted for sharing relations (Geomatics Canada 2002). In Table 8 we have identified the number of errors with respect to the total number of relations. These are within the tolerance regimes set by the CITS, but such errors may cause more important difficulties when data bases from several sources are combined, and hence should be tracked.

Entity-Entity	Relation	Number of objects	Illegal relations
Building- Road	Illegal Superposition	713 - 6204	9
Depot liquid-Depot liquid	Illegal Superposition	110 - 10	1
Contour-Water body	Illegal superposition	2121 - 515	1
Water body-Vegetation	Illegal superposition	515 - 1104	4

Table 4. Inconsistent relations in NTDB.

The overall study of the data file allowed the detection of several types of inconsistencies:

- a) Inconsistent relations: the connection and sharing relations are relatively well respected but problems were found in the superposition and adjacency relations. These are due to the initial conception of the NTDB which represents all the objects as points and lines. The algorithm that reconstructs the polygons from line segments and detects islands in the main polygons seems to fail in some situations.
- b) An important part of the ontology of the NTDB consists of the minimum dimensions for the objects. The areas of some polygonal objects are less than the minimum indicated values in the ontology. There are also many problems concerning the minimum distance allowed between objects of the same kind where the ontology requires a fusion between the objects. Other cases consist of possible connection or sharing relations between objects which are not respected in the data file.
- c) The last category of problems is due to the incomplete ontology. For example, there is no rule that verifies the relation between a road and a building.

The solution for the first two categories must be considered in the control system. For the third category of problem, the lack of certain rules in the ontology must be considered. New rules must be defined in the ontology in order to guarantee the consistency and the completeness of the product ontology.

6. Conclusion

An ontology based method for quality assessment of spatial data is proposed. The overall approach appears to be justified. In the first step the product ontology was translated to Prolog. Then spatial relations in the data level were retrieved using the Geomedia GIS. These relations then were expressed in term of the NTDB relations. In the next step, the formal ontology and the retrieved relations in data level were integrated in a knowledgebase. In the final step, rules that define inconsistencies are developed and the querying of the knowledge base to determine the existence of such inconsistencies was carried out on a very large fact base. The results obtained from various experiments indicate the presence of some inconsistencies in the analyzed data set. These problems are attributed firstly to the control system that oversees the production process and secondly to an incomplete ontology. The process presented in this paper lays the groundwork for subsequent processing steps, such as those concerned with the fusion of ontologies and of the data base itself with other data sources in the broader context of interoperability.

7. Acknowledgments

The authors would like to acknowledge the REVIGIS project and the members of the Canadian Centre for Topographic Information at Sherbrooke for their valuable collaboration.

8. References

- Benferhat, S., Lagrue, S., Papini, O., and Wilson, N., 2002, Representation of imperfect geographic knowledge: (NL), U. Keele (UK), U. Leicester (UK), TU Vienna(AT), CNR/CNUCE (IT), U. Toulon (FR), CRG U. Laval (CA), Praxitec (FR) R212 / N21.
- Bratko, I., 2000, Prolog, programming for artificial intelligence (third ed.), Addison weley, 678 p.
- Egenhofer, M.J., 1993, A model for detailed binary topological relationships: GEOMATICA, v. 47, no. 3&4 Canada, p. 261-273.
- Egenhofer, M.J., and Franzosa, R.D., 1991, Point-set topological relations: International Journal of Geographic Information Systems, v. 5, no. 161-174.
- Egenhofer, M.J., and Herring, J.R., 1990, A mathematical frame work for the definition of topological relations, *in* 4th International Symposium on Spatial Data Handling, Zurich, University of Zurich, p. 803-813.
- Frank, A., 2001, Tiers of ontology and consistency constraints in geographical information systems: International Journal of Geographical Information Science, v. 15, no. 7, p. 667-678.
- Geomatics Canada, 2002, Standards and specifications of NTDB (third ed.): Sherbrooke, 33 p.

- Goodchild, M.F., and Jeansoulin, R., 1998, Data quality in geographic information, from error to uncertainty: Paris, Hermes.
- Guarino, N., and Giaretta, P., 1995, Ontologies and knowledge bases: Towards a terminological clarification, *in* Mars, N., ed., Towards very large knowledge bases: Knowledge building and knowledge sharing: Amsterdam, IOS Press, p. 25-32.
- Guptill, S.C., and Marrison, J.L., 1995, Elements of spatial data quality: Oxford, Elsevier.
- Jeansoulin, R., 2002, Quality of geographic information: Ontological approach and artificial intelligence tools in the revigis project, *in* EC-GI& GIS Workshop, p. 12.
- Mark, D., Egenhofer, M.J., and Hirtle, S., Smith, B., 2001, Ontological foundation of geographic information science: UCGIS Emerging Research Theme.
- Mostafavi, M.A., Edwards, G., Jeansoulin, R., 2002, Ontological consistency of spatial databases in support of data fusion. Colloque Geomatics 2002, Montréal, Canada.
- Shi, W., Goodchild, M.F., and Fisher, P.F., 2002, Spatial data quality (1st ed.), Taylor & Francis, 340 p.
- Smith, B., and Mark, D., 2001, Geographical categories: An ontological investigation: International Journal of Geographical Information Science, v. 15, no. 7, p. 591-612.
- Veregin, H., and Hargitai, P., 1995, An evaluation matrix for geographical data quality., *in* Morrison, J.L., ed., Elements of spatial data quality.: Oxford, Elsevier, p. 167-188.